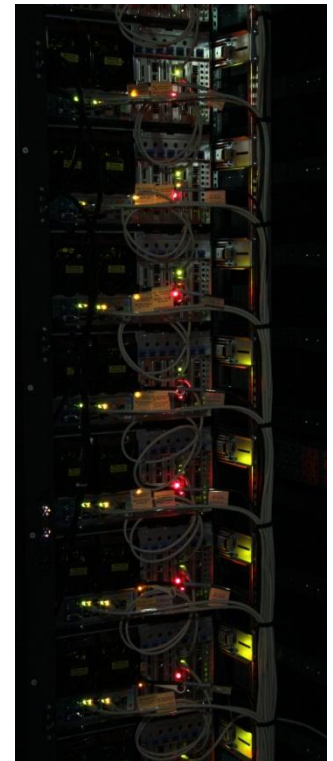


RWTH GPU Cluster



Fotos: Christian Iwainsky

Sandra Wienke
wienke@rz.rwth-aachen.de
March 2013





GPU Cluster: 57 Nvidia Quadro 6000 (Fermi)

“innovative computer architecture”

■ High utilization of resources

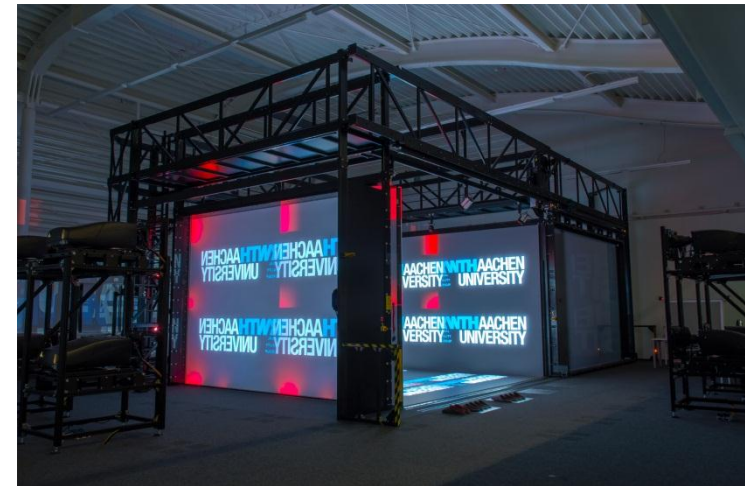
→ Daytime

→ VR: new CAVE (49 GPUs)

→ HPC: interactive software development
(8 GPUs)

→ Nighttime

→ HPC: Processing of GPGPU compute jobs (55-57 GPUs)



aixCAVE, VR, RWTH Aachen, since June 2012

		4	24	1
		dialogue nodes	rendering nodes	head node
Name		linuxgpud[1-4]	linuxgpud[01-24]	linuxgpud1
Devices	#	2		1
	details/GPU	<p>NVIDIA Quadro 6000 (Fermi) 448 cores 1.15 GHz 6 GB RAM ECC on max. GFlops: 1030.4 (SP), 515.2 (DP)</p>		
Host	processor	<p>2 x Intel Xeon X5650 EP (Westmere) (12-core CPU) @ 2.67GHz</p>		
	RAM	24 GB		48 GB
Network		QDR InfiniBand		

Software stack

- Environment as on compute cluster (modules,...)

- **CUDA Toolkit: 5.0 (4.1, 4.0, 3.2)**

- CUDA

- OpenCL (1.0)

```
module load cuda
→ directory: $CUDA_ROOT
```

- **PGI Compiler**

- CUDA Fortran

- PGI Accelerator Model

- PGI OpenACC

```
module load pgi
(module switch intel pgi)
```

- **Intel OpenCL SDK:** OpenCL (1.1) for Intel CPUs

```
module load intelopencl
```

- **CUDA Debugging**

- TotalView

```
module load totalview
```

- DDT

```
module load ddt
```

■ GPUs + Windows

- ½ NVIDIA Tesla S1070 (2 GT200 GPUs)
- Host: 8-core Intel X5570 (Nehalem) @ 2.93 GHz
- Interactive + batch mode
- Software: CUDA Toolkit, Matlab, Parallel Nsight Debugger,...



■ Getting access to GPU systems

- E-mail to servicedesk@rz.rwth-aachen.de
 - OS: Linux GPU-Cluster and/or single Windows machines?
 - Short description of your (GPU) application (or your purposes)
 - Programming paradigm (e.g. CUDA, OpenCL,...)?
 - Single or multi GPU usage?
- Includes access to GPGPU-Wiki (full documentation)

Lab: Login to GPU machines

- **User name (see handout):** `hpc1ab<XY>`
- **Password (see handout)**
- **GPU node (see next slide):** `linuxgpus<AB>`
- **Jump from frontend node to GPU node:** `ssh -Y linuxgpus<AB>`

- **Remark: GPUs are set to “exclusive mode” (per process)**
 - Only one person can access GPU
 - If occupied, e.g. message “all CUDA-capable devices are busy or unavailable”

These nodes are available in interactive mode only for the workshop!

See what is running: nvidia-smi

```
linuxgpud1$> nvidia-smi  
Mon Oct 17 12:41:01 2011
```

nvidia-smi -q
Lists GPU details

ECC (SB: single bit,
DB: double bit)

```
+-----+  
| NVIDIA-SMI 2.285.05   Driver Version: 285.05.09   |  
+-----+  
| Nb. Name GPU ID + type | Bus Id      Disp. | Volatile ECC SB / DB |  
| Fan  Temp  Power Usage /Cap | Memory Usage | GPU Util. Compute M. |  
+-----+  
| 0.  Quadro 6000 | 0000:02:00.0 Off | 0 0 |  
| 30%  80 C  P0   Off / Off | 4% 208MB / 5375MB | 99% E. Process |  
+-----+  
| 1.  Quadro 6000 | 0000:85:00.0 On  | 0 0 |  
| 36%  84 C  P8   Off / Off | 0% 22MB / 5375MB | 0%  E. Process |  
+-----+  
| Compute processes: |  
| GPU  PID  Process name | usage |  
+-----+  
| 0.  30234 nbody | 196MB |  
+-----+
```

GPU ID + type

ECC SB / DB

E. Process

E. Process

compute mode: 1 person (1 process)

process running on GPU